Amendments to the Specification:

1. Please perform all of the amendments indicated in the Applicant's October 20, 2003 response to the Office Action of April 22, 2003, except amendment No. 8, which contained an inadvertent text transposition, and which is corrected in the following amendment No. 2.

2. Please replace the text in Applicant's October 20, 2003 Response to the Office Action of April 22, 2003, in the Amendment to the Specification No. 8:

> 8. Please delete the text beginning on page 9, line 12:
> genes in the clusters. The figureof-merit indices of the method and Indices are then generated by testing the model's ability to classify additional text about system relate to the percentage of times that the tested classifications are made correctly, as compared with classifications performed on text corresponding to genes placed randomly in clusters.

with the following amended text:

> 8. Please delete the text beginning on page 9, line 12:
> Indices are then generated by testing the model's ability to classify additional text about genes in the clusters. The figure-of-merit indices of the method and system relate to the percentage of times that the tested classifications are made correctly, as compared with classifications performed on text corresponding to genes placed randomly in clusters.

3. Please replace the text on page 4, line 12:

> http://www-stat.stanford.edu/~tibs/lab/publications.html

with the following amended text:

> www-stat.stanford.edu/~tibs/lab/publications.html

4. Please replace the text on page 7, line 4:

> http://www.cs.washington.edu/homes/kayee/research.html

with the following amended text:

www.cs.washington.edu/homes/kayee/research.html

5. Please replace the text on page 16, line 20:

http://www.fsf.org

with the following amended text:

www.fsf.org

6. Please replace the text on page 16, line 22:

ftp://ftp.simtel.net/pub/simtelnet/gnu/djgpp

with the following amended text:

ftp.simtel.net/pub/simtelnet/gnu/djgpp

7. Please replace the text on page 17, line 7:

http://www.cs.cmu.edu/~mccallum/bow

with the following amended text:

www.cs.cmu.edu/~mccallum/bow

8. Please replace the text on page 18, line 14:

http://www.cs.cmu.edu/~mccallum/bow

with the following amended text:

www.cs.cmu.edu/~mccallum/bow

9. Please replace the text on page 24, line 3:

http://www.ncbi.nlm.nih.gov/htbin-post/Omim/dispmim?"Omimnumber"

with the following amended text:

www.ncbi.nlm.nih.gov/htbin-post/Omim/dispmim?"Omimnumber"

10. Please replace the text on page 25, line 11:

http://www.ncbi.nlm.nih.gov/htbin-

post/Entrez/query?db=m&form=6&dopt=l&html=no&uid=UID

with the following amended text:

www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?db=m&form=6&dopt=l&html=no&uid=UID

11. Please replace the text on page 25, line 17:

http://www.ncbi.nlm.nih.gov/entrez/utils/qmap.cgi

with the following amended text:

www.ncbi.nlm.nih.gov/entrez/utils/qmap.cgi

12. Please replace the text on page 25, line 18:

http://www.ncbi.nlm.nih.gov/entrez/utils/qmap_help.html

with the following amended text:

www.ncbi.nlm.nih.gov/entrez/utils/qmap_help.html

13. Please replace the text on page 29, line 8:

http://lib.stat.cmu.edu/general/clusfind

with the following amended text:

lib.stat.cmu.edu/general/clusfind

14. Please replace the text on page 51, line 2:

http://genome-www.stanford.edu

with the following amended text:

genome-www.stanford.edu

15. Please replace the text on page 51, line 12:

http://genome-www.stanford.edu

with the following amended text:

genome-www.stanford.edu

16. Please replace the text on page 36, line 9:

Text Modeling and Classification

with the following amended text:

Text Modeling

17. Please replace the text on page 36, line 14:

It then produces a statistical model of the text that is suitable for text classification.

with the following amended text:

It then produces a statistical model of the text that is suitable for text classification, although it should be noted that the disclosed invention does not actually perform text classification and does not use the features of Rainbow that actually perform text classification.

18. Please replace the text on page 38, line 7

The information that is provided automatically by the Keyword Identification Module **(128)** is a list of words for each cluster, sorted in descending order according to the numerical weights calculated by a classification algorithm.

with the following amended text:

The information that is provided automatically by the Keyword Identification Module **(128)** is a list of words for each cluster, sorted in descending order according to the numerical weights calculated by a classification algorithm. (It should be noted that only a portion of the algorithm is used to provide the list of words for each cluster, and that portion does not actually perform text classification. The relevant portion of the algorithm is instead a preliminary or adjunct to the portion that would perform text classification).

19. Please insert the following text, beginning as a new paragraph, after the text ending on page 10, line 15:

There is little prior art that that can assist an artisan in automatically generating a useful corpus of literature about an individual gene, which might then be used to analyze the literature about the genes that constitute a microarray cluster. PubMed/MEDLINE is the most widely used on-line source for gene related abstracts and literature, which might be used to generate such a corpus, but few investigators have described its use for any similar purpose. SHATKAY et al (2000) explain that PubMed provides for literature search and retrieval by two methods -- boolean query and similarity query (also known as "neighboring"). They describe how there are well-known deficiencies with any attempt to use the method of boolean queries to generate a text corpus. For example, CHAUSSABEL and SHER (2002) attempted to use boolean queries consisting of gene names taken from a list, and ultimately found it necessary to manually edit or correct the unacceptably large number of errors that resulted from use of boolean queries. Accordingly, Shatkay et al. advocate using only the neighboring feature of PubMed to acquire a set of documents about a gene, after first selecting a "kernel" citation for that gene (if possible) within a curated database about the genes under investigation. The literature in PubMed that "neighbors" this kernel citation is then generated by PubMed after providing it the kernel citation as the neighboring query. The method of Shatkay et al then seeks to find similarities within the documents so generated for different genes. This method can be automatic only if there already exists a curated citation list from which to obtain the "kernel" documents, as was the case with the yeast genes investigated by Shatkay et al. Otherwise, and in general, an expert human would need to select the kernel documents. Furthermore, Shatkay et al. teach

that when a clustering of genes is already available from microarray expression experiments, then that clustering should be ignored, except for purposes of manually comparing with results obtained independently by their method. Another method for generating a corpus of text using MEDLINE was described by ANDRADE and VALENCIA (1998), but it was used to generate a corpus only for protein domain families, rather than for individual, arbitrarily selected genes. According to their method, protein families in the PDBSELECT database pointed to entries in the SwissProt database, which pointed to articles in MEDLINE, which were then taken to be the literature corpus for the corresponding protein domain family. This method is not generally applicable to the problem of generating a literature corpus for an arbitrarily selected gene, because a gene may not belong to a known protein family. Furthermore, the size of the literature corpus would be limited by the number of pointers in the SwissProt database. Given the above-mentioned limitations of the prior art, it was therefore an aim of the present invention to provide an original method for automatically generating a substantial literature text corpus for an arbitrarily selected gene, which could be used to generate a literature text corpus for clusters obtained from microarray experiments.